CLAIMS:

1. A method for normalizing input strings, the method comprising the steps of:

(a) receiving the input strings;

(b) linguistically analyzing the input strings to generate a first representation of each of the input strings; each of the first representations including linguistic information;

(c) skeletising each of the first representations to generate a corresponding second representation for each of the input strings; said skeletising step replacing the linguistic information with abstract variables in each of the second representations; and

(d) storing the second representation as normalized representations of the input strings.

2. The method of claim 1, wherein said step of linguistically analyzing comprises performing a plurality of operating functions.

3. The method of claim 2, wherein said plurality of operating functions comprise performing one of morphological analysis, syntactic analysis, and semantic analysis.

4. The method of claim 3, wherein said step of linguistically analyzing comprises normalizing words according to their base forms.

5. The method of claim 3, wherein said analysis further comprises the step of extracting a syntactic category for individual words.

6. The method of claim 3, wherein said analysis further comprises the step of extracting syntactic information representing string structure.

7. The method of claim 3, wherein said analysis further comprises the step of extracting dependency relations between sub-structures of a string.

8. The method of claim 3, wherein said analysis further comprises providing semantic links for individual words.

9. The method of claim 2, further comprising the step of performing machine learning for selecting particular operating functions out of said plurality of operating functions and for determining the processing order.

9. The method of claim 2, wherein said storing further comprises storing the operating functions performed on the normalized representations.

10. The method of claim 1, wherein the abstract variable are tags indicating the replaced linguistic information.

11. The method of claim 1, wherein the normalized representations are stored in a database.

12. The method of claim 11, further comprising:

receiving a query;

generating a normalized representation of said query by performing steps (b) and (c);

matching the normalized representation of said query to the normalized representations stored in the database; and

retrieving from said database strings identified by said matching step.

13. The method of claim 1, wherein said steps (a) – (d) are performed to generate a translation memory comprising a plurality of normalized representations of strings in a first language and a second language.

14. The method of claim 13, further comprising the steps of:

receiving an input string in the first language;

retrieving a similar string in said first language from said plurality of normalized representations, and

outputting said translation information based on a string in said second language which corresponds to said retrieved string in said first language.

15. An apparatus for normalizing input strings, the apparatus comprising:

a text processing unit for:

receiving the input strings,

linguistically analyzing the input strings to generate a first representation of each of the input strings; each of the first representations including linguistic information, and

skeletising each of the first representations to generate a corresponding second representation for each of the input strings; said skeletising replacing the linguistic information with abstract variables in each of the second representations; and

memory for storing the second representation as normalized representations of the input strings.

16. The apparatus of claim 15, further comprising a query formatting unit for:

receiving queries;

linguistically analyzing the queries to generate a first representation of each of the queries; each of the first representations including linguistic information; and

skeletising each of the first representations to generate a corresponding second representation for each of the queries; said skeletising replacing the linguistic information with abstract variables in each of the second representations.

17. The apparatus of claim 16, further comprising:

memory for storing the second representation of the queries as normalized representations of the queries;

a matching unit for matching the normalized representations of the input strings with the normalized representation of the queries.

18. The apparatus of claim 16, further comprising a translation memory for storing translations of the input strings.